PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : G09B 19/04

A1

(11) International Publication Number: **WO 99/13446**

(43) International Publication Date: 18 March 1999 (18.03.99)

(21) International Application Number: PCT/IL98/00426

(22) International Filing Date: 2 September 1998 (02.09.98)

(30) Priority Data:
08/923,955 5 September 1997 (05.09.97) US

(71) Applicant *(for all designated States except US)*: IDIOMA LTD. [IL/IL]; Yozmot Granot Initiative Center, 38100 D.N. Hefer (IL).

(72) Inventor; and
(75) Inventor/Applicant *(for US only)*: GOTTESFELD, Ziv [IL/IL]; Derech Hayam 9, 40293 Beit Yannai (IL).

(74) Agent: BEN-DAVID, Yirmiyahu, M.; Jeremy M. Ben-David & Co., Har Hotzvim Hi-Tech Park, P.O. Box 45087, 91450 Jerusalem (IL).

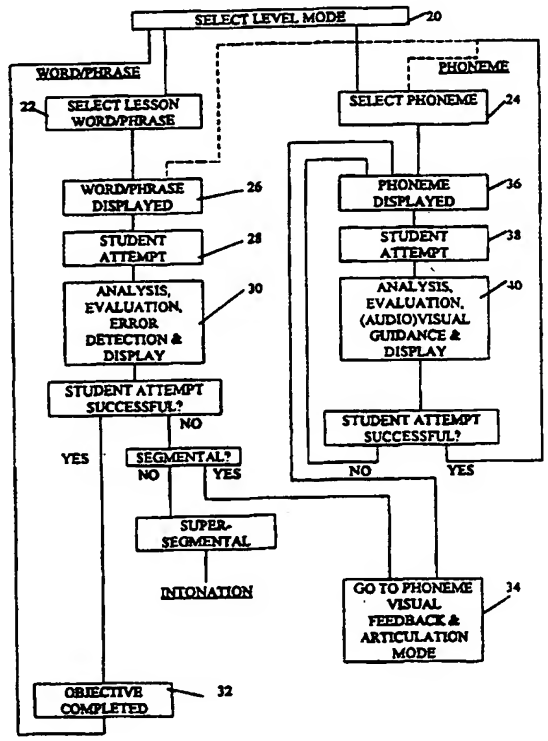
(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, HR, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: INTERACTIVE SYSTEM FOR TEACHING SPEECH PRONUNCIATION AND READING

(57) Abstract

A system for teaching speech pronunciation or reading to a student, has a memory (10) for storing a plurality of speech portions; a playback system (112, 18), associated with the memory, for indicating to a student (USER) a speech portion to be practiced; an algorithm associated with the memory and the playback system; a speech portion selector (20) for selecting a speech task (22, 24) to be practiced; and a sound recorder apparatus (16) operatively connected to the algorithm and operative to sense and record a sound uttered by a student, and to provide the utterance to the algorithm in signal form, wherein the speech recognition algorithm (30, 40) is operative to compare the utterance with the speech portion to be practiced and to evaluate the accuracy of the utterance; and the playback system being operable provide to the student an indication of the accuracy of the utterance (64, 68).



FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania
AM	Armenia
AT	Austria
AU	Australia
AZ	Azerbaijan
BA	Bosnia and Herzegovina
BB	Barbados
BE	Belgium
BF	Burkina Faso
BG	Bulgaria
BJ	Benin
BR	Brazil
BY	Belarus
CA	Canada
CF	Central African Republic
CG	Congo
CH	Switzerland
CI	Côte d'Ivoire
CM	Cameroon
CN	China
CU	Cuba
CZ	Czech Republic
DE	Germany
DK	Denmark
EE	Estonia
ES	Spain
FI	Finland
FR	France
GA	Gabon
GB	United Kingdom
GE	Georgia
GH	Ghana
GN	Guinea
GR	Greece
HU	Hungary
IE	Ireland
IL	Israel
IS	Iceland
IT	Italy
JP	Japan
KE	Kenya
KG	Kyrgyzstan
KP	Democratic People's Republic of Korea
KR	Republic of Korea
KZ	Kazakstan
LC	Saint Lucia
LI	Liechtenstein
LK	Sri Lanka
LR	Liberia
LS	Lesotho
LT	Lithuania
LU	Luxembourg
LV	Latvia
MC	Monaco
MD	Republic of Moldova
MG	Madagascar
MK	The former Yugoslav Republic of Macedonia
ML	Mali
MN	Mongolia
MR	Mauritania
MW	Malawi
MX	Mexico
NE	Niger
NL	Netherlands
NO	Norway
NZ	New Zealand
PL	Poland
PT	Portugal
RO	Romania
RU	Russian Federation
SD	Sudan
SE	Sweden
SG	Singapore
SI	Slovenia
SK	Slovakia
SN	Senegal
SZ	Swaziland
TD	Chad
TG	Togo
TJ	Tajikistan
TM	Turkmenistan
TR	Turkey
TT	Trinidad and Tobago
UA	Ukraine
UG	Uganda
US	United States of America
UZ	Uzbekistan
VN	Viet Nam
YU	Yugoslavia
ZW	Zimbabwe

INTERACTIVE SYSTEM FOR TEACHING SPEECH PRONUNCIATION & READING

FIELD OF THE INVENTION

The present invention relates to speech teaching, generally, and, in particular, to a software-based system for the teaching of correct speech and reading.

BACKGROUND OF THE INVENTION

It has long been sought to provide ways of teaching correct pronunciation of a particular language, inter alia, for the purpose of learning correct pronunciation of a foreign language, for general speech therapy. It has also been sought to provide a tool for evaluation and guidance of one's reading of a native or foreign language.

Known art includes the following patent publications:

US Patent No. 4,636,173 which discloses a method of teaching reading by synchronizing a visual display with a soundtrack by momentarily highlighting displayed words as they are emitted by the soundtrack.

US Patent No. 5,142,657 which relates to a computerized system for providing a visual output of analyses of speech including the parameters of waveform, power, pitch and sound spectrograph, and comparing these parameters with corresponding model parameters.

US Patent No. 5,286,205, which relates to a method of teaching spoken English using mouth position characters. This method is based on a visual display of mouth positions required for different pronunciations.

US Patent No. 5,393,236 which describes a computer-based interactive speech pronunciation apparatus and method.

US Patent No. 5487671 relates to a computerized system for teaching speech that evaluates accuracy of pronunciation relative to a stored database according to one or more speech parameters.

US Patent No. 5,503,560 which relates to a computerized system for speech training, in which a user is prompted in the pronunciation of keywords. The system records a first attempt at pronunciation of a keyword and compares subsequent attempts with the first attempt. An improvement in pronunciation is claimed to be correlated with a significant deviation in user's speech template. There is also provided a display which shows a required mouth shape for the sounds to be learned. A video analysis of the

user's actual mouth positions may also be provided by use of a video pick-up and analyzer.

Published PCT application no. WO 91/00582 which relates to a system which compares pronunciation of a word or sentence with a reference word or sentence, and which provides audio and video displays of the comparison.

The above patent publications are characterized by various disadvantages, as follows:

US Patent No. 4,636,173 discloses a method which is not interactive, and thus does not provide any indication to a student as to the accuracy of his pronunciation, nor does it indicate a way of achieving correct pronunciation.

US Patent No. 5,142,657 provides a computerized method which does not provide an easily interpretable feedback, and does not provide an explanation of how to improve pronunciation, merely which parameters of speech need to be improved. Furthermore, a display of these parameters, while they may be suitable for expert users in language laboratory, will not be helpful to less skilled students or children.

US Patent No. 5,286,205 relates to a teaching method which is not interactive, such that a student has to judge for himself whether or not his pronunciation is correct, there being no objective feedback thereof. Furthermore, the method teaches use of different mouth positions, and cannot therefore be used for all sounds for which the mouth position is not the only important key to correct pronunciation.

US Patent No. 5,393,236 relates to a computer-based interactive speech pronunciation method which is not self-sufficient and which requires supervision by an instructor and, moreover, does not in any way guide user towards correct pronunciation.

US Patent No. 5487671 relates to a computerized system for teaching speech that evaluates accuracy of production relative to a stored database according to one or more speech parameters. It does not provide to the user an indication as to how to improve his pronunciation, nor does it point to the user the nature of his mistakes, nor does it provide an algorithm that allows the evaluation of a given pronunciation, nor does it provide a methodology for dealing with pronunciation mistakes at various levels.

US Patent No. 5,503,560 relates to a computerized system that judges improvement in pronunciation according to a deviation in user's own voice, but it does not directly compare user's pronunciation to that of native speakers of the language, nor does it direct the user as to how to improve his pronunciation, nor does it point out to the user the mistakes made within the phrase or keyword.

Published PCT application no. WO 91/00582 describes a system which does not provide any indication of how to achieve correct pronunciation.

In general, known methods do not enable a student either to learn correct pronunciation of parts of speech or to learn how to read, such as when the student is a child being taught to read in his native language, wherein the feedback is based on totally objective criteria, and is totally interpretable by and thus immediately useful to a student without requiring interpretation or guidance by an instructor.

SUMMARY OF THE INVENTION

It is thus an aim of the present invention to provide a fully interactive, self-contained system for teaching pronunciation of language sounds. This system may also be used for teaching a person how to read in his native language. In particular, a speech recognition algorithm is provided so as to enable full interaction between a student and the system, in real time.

In particular, the software of the invention is employed in the system such that, in response to selected utterances, one or more visual stimuli, such as one or more moving images on a visual display unit are activated in a desired manner. An utterance which is not sufficiently accurate activates the stimulus, but not in the desired manner. Instruction is provided by way of displaying the correct tongue position inside the mouth, also known as articulatory positioning.

As will be appreciated from the description hereinbelow, the system of the invention operates both at the level of the individual phoneme, and also at the level of multi-phoneme strings, such as words and phrases.

There is thus provided, in accordance with a preferred embodiment of the invention, a system for teaching speech pronunciation or reading to a student. The system includes a memory for storing a plurality of speech portions, a playback system, associated with the memory, for indicating to a student a speech portion to be practiced, and a speech portion selector for selecting a speech task to be practiced.

The system is operated via an algorithm which is associated with the memory, the playback system, and also with a sound recorder which is operative to sense and record a sound uttered by a student, and to provide the utterance for processing by use of the algorithm, in signal form. The algorithm performs a comparison of the utterance with the speech portion to be practiced, evaluates the accuracy of the utterance, and provides

real time, an indication of the accuracy of the utterance.

Further in accordance with a preferred embodiment of the present invention, the speech portion selector includes apparatus for selecting a phoneme in a selected phoneme class, and the algorithm is also operative to determine whether or not a phoneme present in the utterance belongs to the selected phoneme class. If a phoneme in the utterance is determined to be outside a selected phoneme class, then the utterance is 'rejected' as being inaccurate, and the student may be instructed to try again, if the system is operating at the single phoneme level. If the system is operating in the multi-phoneme string or word/phrase level, the student may be informed of the problematic phoneme or phonemes, referred to also herein as "subgroups" , and instructed to practice them before proceeding with the more complex task.

Additionally in accordance with a preferred embodiment of the present invention, the playback system includes visual playback apparatus and audio playback apparatus, and, in response to selection of a selected speech portion by a student, the visual playback apparatus is operative to display a visual image indicating the speech portion selected, and the audio display apparatus is operative to provide an audible indication of the speech portion selected.

Further in accordance with a preferred embodiment of the present invention, the playback system is operative to provide, preferably in real time, a dynamic visual image indicating the accuracy of the utterance.

Additionally in accordance with a preferred embodiment of the present invention, the playback system includes apparatus for displaying a movable visual image which is movable between first and second locations on the display, wherein the first location is a start location at which the visual image is located prior to sensing of a sound by the sound recorder, and wherein the second location is a target location, towards which the playback system is operative to move the movable visual image in real time as an indication of the accuracy of the utterance.

Further in accordance with a preferred embodiment of the present invention, the distance between the movable visual image and the target location is inversely proportional to the accuracy of the utterance.

BRIEF DESCRIPTION OF THE DRAWINGS

The present invention will be more fully understood and appreciated form the following detailed description, taken in conjunction with th drawings, in which:

Fig. 1 is a block diagram representation of an interactive speech pronunciation teaching system, constructed and operative in accordance with the present invention;

Fig. 2 is a schematic representation of the sequence of events employed in the present invention to teach correct speech and pronunciation;

Fig. 3A is a diagrammatic representation of a visual display "prompt" screen provided by the system to a student, in response to selection by the student of a particular word or phoneme;

Fig. 3B is a diagrammatic representation of a prompt screen provided by the system, in response to an incorrect pronunciation of a subgroup within a multi-phoneme string, visually emphasizing the incorrectly pronounced subgroup;

Fig. 3C is a diagrammatic representation of a real time, visual feedback prompt screen; and

Fig. 4 is a flow chart of the methodology employed by the present invention to analyze the speech of a student and to provide visual feedback of the student's performance.

DETAILED DESCRIPTION OF THE INVENTION

The present invention seeks to provide a computerized system for the teaching of correct speech, particularly by use of speech recognition algorithms. The object of the system is the interactive teaching of correct speech in a language, either foreign or native, as well as for speech therapy. The system may also be used to teach reading, particularly to a child, of his native language. As will be appreciated from the following description, the system utilizes a variety of techniques in order to achieve this task, including: speech recognition, speech evaluation; speech error detection; accent recognition; real time visual feedback; audio feedback; and articulatory guidance.

The system may be adapted for use with a variety of computer systems, and may, in a preferred embodiment of the present invention, be fully self-contained within, for example, a suitable multimedia personal computer. In accordance with other embodiments of the invention, the present teaching system may be adapted for use via any suitable multimedia-enabled computerized platform which may or may not be constructed specifically for the system of the invention, or the system may be based on a computer network, such as Internet, intranets and the like.

Referring now to Fig. 1, the system based on a preferred embodiment of the present invention includes a computer 10 equipped with a sound card, a visual display

unit (VDU) 12, typically, a high-speed color monitor, a manual data input unit 14, which may be a keyboard and/or a pointing device, such as a mouse, glide pad, or a touch-sensitive screen forming part of VDU 12, a microphone 16, and a speaker 18. The hardware components shown and described herein may be totally conventional, and thus, no further specific description thereof is necessary.

Referring now to Fig. 2, there is shown, a schematic representation of the sequence of events in a typical speech and pronunciation teaching session with a student, according to a preferred embodiment of the present invention. Examples of "prompt" screens displayed to the student in such a session are shown in Figs. 3A through 3C.

In a system implemented according to a preferred embodiment of the present invention, a typical speech and pronunciation teaching session includes the following sequence of events, shown schematically in Fig. 2:

First (block 20): the student selects a "level" mode, typically including either a lesson containing plural phonemes, such as a word or phrase (block 22), or subgroups of such, or a single phoneme (block 24).

It will be understood that a 'lesson' is a collection of production tasks having a common denominator. For example, in a production lesson for the phoneme /l/, the lesson plan would be words and phrases containing /l/ in various positions in the lesson words.

Second (block 26): in the event that that word or phrase level has been selected, one or more words or sentences containing the lesson subject, exemplified as the word "SHELL," shown on prompt screen 60 in Fig. 3A, are presented to the student. Preferably, the word is also sounded by the system, so that the student hears the correct pronunciation, which he is then to repeat.

In accordance with an alternative embodiment of the invention, the system may also be taught how to pronounce specific words or sounds, for example, so as to adapt it to a particular regional accent. In this case however, various default accents are retained in the system's memory so as to enable the system to be reset, if required.

Third (block 28): the student repeats the lesson word or words into the microphone 16 (Fig. 1).

Fourth (block 30) : the student's speech is analyzed and evaluated for errors in pronunciation; errors are indicated as by a display prompt, as seen on prompt screen 62 in Fig. 3B, in which the subgroup, in this case the phoneme /l/, is indicated as having been mispronounced.

In the event that the student pronunciation was successful, such that the objective has been completed (block 32) (by a correct pronunciation of the selected lesson subject), he may then be returned either to the level select mode (block 20), or to the lesson select mode (block 22).

In the event that the student pronunciation was not successful, it is determined whether his mispronunciation is "segmental," namely, relating to the phrase/word/phoneme levels or, "super-segmental." If the problem is super-segmental, i.e. it relates to stress or intonation, he is referred to a system dealing with that particular problem. That type of system is known in the art, and is thus beyond the scope of the invention, and is thus not dealt with herein. If the problem is determined to be segmental, then the student is transferred to the phoneme visual feedback and articulation mode (block 34), where he has the option of studying the inaccurately pronounced subgroup not only by imitation, but also by being shown the correct articulation, or required tongue positioning. In this mode, the system points out to the user the nature and location of his mistake.

Accordingly, at this stage, the system provides the student with the option of replaying own audio recording, while at the same time, providing a visual display of the subject phoneme (block 36). The system may also replay a model audio recording, for purposes of comparison.

The student then attempts to repeat the subgroup (block 38), which the system analyses and evaluates (block 40). If the student is unable to improve performance, the system enters visual feedback mode, indicated as "(audio)visual guidance and display" in block 40, which is shown and described herein in conjunction with Fig. 3C below.

Once the student has improved performance in this mode, the system may return either to the word/phrase display level (block 26), or to the phoneme select level (block 24), if he decides that single phonemes should be practiced and acquired prior to proceeding to word or phrase (subgroup) lessons. Otherwise, he is returned to a level whereat he practices the phoneme with which he is having trouble.

By way of example, consider a case wherein the subject of the lesson is correct pronunciation of the /l/ sound. The student is shown an animation of the word "SHELL" 60 on the visual display unit 12 (Fig. 1), as well as an appropriate icon (not shown) representing a shell. The system then plays back a model recording of "SHELL", and prompts the student to repeat it into the microphone 16. The student is provided with both visual and audio prompts.

When the student repeats the word, and, for example, mispronounces /l/, the system points out the error to the student, seen at 64 in FIG. 3B. This will be by some form of animation (not shown), as well as by an audio indication such as "you have mispronounced the /l/ in "shell."

The student can choose either to try to pronounce the word again correctly, or to receive further guidance in the form of real time visual feedback and articulation.

Real time visual feedback is based on the student's control of a targeting device appearing as on a prompt screen 66, seen in Fig. 3C. By use of a speech recognition algorithm, as described below, the system extracts predetermined relevant speech parameters from the student's rendition of a test phrase, word or phoneme, and transforms the student's performance into a distance from the appropriate target. In the example shown in Fig. 3C, the student is shown a prompt screen with /l/ as a target, the other target being the phoneme /r/. Different targets may also be provided, a 'default' target being the relevant 'mistake.' In other words, if a common mistake made when pronouncing /l/ is the phoneme /r/, then the default target, as seen in the drawings, will be /r/.

There exists, however, the option, particularly when the system of the invention is used within a supervised setting, of a supervisor (clinician or teacher) adding, changing or removing 'target' points of reference.

A targeting device, referenced 68, is also shown, being exemplified by a circle, which is initially positioned at a 'zero' position, over a pair of cross hairs 70 and 72.

Each time the student repeats the phoneme /l/, the targeting device 68 moves closer to or further from the target phoneme /l/, wherein the displayed distance between the targeting device 68 and the target phoneme is inversely proportional to the perceived acoustic "distance," or accuracy of the pronounced phoneme. If the student pronounces the phoneme correctly, the targeting device 68 is moved into coincidence with the target phoneme. An animation or other entertaining event may also be shown by way of reward.

In accordance with the present invention, a “correct” pronunciation is that whose extracted speech parameters are substantially the same as those of a database of recordings of that single sound, word or phrase, (which may also be referred to as a “multiple phoneme string”), adjudged to be well pronounced by a group of experts, such as speech therapists, or professional teachers of the language.

In accordance with an alternative embodiment of the invention, however, the system may be 'taught' or adjusted online so as to a new definition of 'correct.' For

example, if a particular pronunciation is perceived to be good, even though the system judged it as "bad", the system may be adjusted so as to accept that particular sound as valid or correct, either for a particular user, or, in general, for a group of users.

Additional options of articulatory guidance available to the student are graphical and acoustic demonstrations. A further option allows the student to receive visual feedback in the form of spectrum and spectrographic real time display, with acoustic targets superimposed, (not shown).

In a preferred embodiment of the present invention, the student is guided towards correct pronunciation by real time visual feedback. The analysis required to convert the student's rendition of the test phoneme or word is shown schematically in Fig. 4 and includes the following steps, all of which are performed in real time, by use of appropriate algorithms, as described below.

It will be appreciated by persons skilled in the art that the system of the present invention is operative to enable the provision of feedback to a user, in real time, due to the use of novel speech recognition algorithms, as described below in conjunction with Fig. 4.

While the speech algorithms and portions of the technique or techniques described below are known in various different fields, the use of speech recognition software in order to provide real time, objective speech pronunciation instruction, at the phoneme/word/phrase levels, such as in the present invention, is not known, per se, nor is it believed to have been considered in the art.

In particular, the following techniques are provided in the invention, and are described in detail hereinbelow, namely:

1. Primary Filtration: extraction of features enabling initial exclusion of fundamentally incorrect sounds.
2. Statistical Analysis: filtering procedures for enhancing the relevant parameters, exemplified herein as cepstral parameters, and for reducing the weight of those which are not.
3. Secondary Filtration: The use of "clustering pronunciation filters," for filtering out mispronunciations.
4. Continuous Classification Network: Determining location of sounds in relevant phonetic space.

Primary Filtration

As a prerequisite to performing the analysis of the invention, a database of "correctly pronounced" phonemes and words is collected. As described above, this may be changed for the needs of a particular user.

Accordingly, when detecting a sound, the speech parameters or features thereof are extracted (block 44) by the system by using "cepstral" techniques, as described in the book entitled "Discrete Time Processing Of Speech Signals," by John R. Deller, John G. Proakis, & John H. Hansen, published By MACMILLAN PUBLISHING CO., NY, 1993. This includes calculation of the cepstrum, 1st and 2nd cepstral derivatives, determination of pitch, energy and zero crossing (i.e. the number of times in a given time period that the speech signal crosses a zero level so as to switch between positive and negative values and vice versa). These data are used so as to enable a primary filtration of fundamentally mispronounced sounds, i.e. those which are adjudged to be out of bounds of the defined task.

If the detected sounds are not rejected based on the above primary filtration, they are then subjected to a statistical analysis (block 46), prior to being passed to a secondary pronunciation filter (block 48).

Statistical Analysis

The statistical analysis includes two main steps and is used to determine the number of, and the nature of, the most relevant parameters, and to reduce the dimensionality of the system of parameters.

The first step is in applying previously determined statistical weighting functions, thereby to enhance those parameters most relevant to the particular task at hand. These parameters are those which are statistically predetermined to have greater relevance to the task at hand.

Subsequently, in a second step, all of the above parameters, regardless of weighting, are analyzed by use of Principal Component analysis, also known in the art as the Karhunen-Loeve Transform. This analysis provides a new set of parameters, each being a linear combination of the previous, weighted parameters, such that, the ranking of the new parameters is a function of the variability and thus also of task relevance thereof. After obtaining the new set of ranked, weighted parameters, the parameters set can be truncated so as to reduce dimensionality thereof, while retaining a number of parameters which has been predetermined to be statistically representative of the task data.

Secondary Filtration

As known in the art, phonemes can be grouped into major classes which share specific features. The secondary filtration stage includes performing of a geometric cluster analysis, in order to filter out utterances of individual phonemes that fall outside the class to which the particular acoustic production task relates. For example, if the task were the correct utterance of various sounds in a particular fricative class, such as /s/, /f/, and so on, any mispronounced sounds which, by definition, could not be placed in the same phoneme class as these aforementioned fricatives, such as a "lateral" /s/, or a /z/, would be filtered out or rejected.

Continuous Classification Network

Subsequently, a non-linear, continuous, classification network (block 50), based on such methods as neural network, radial basis function (RBF) sets, or other, is trained using the extracted parameters so that it's output will continuously span the relevant "phonetic space." The continuous classification network is employed in order to determine where exactly the detected sound resides within the relevant phonetic space. Referring to the last example, the detected sound, passing the cluster analysis based filter, may now be detected to reside anywhere between the /s/ and /f/ sound. The targeting device will then be positioned accordingly.

If the system is being operated in the word/phrase level mode, such that the sounds spoken by the student, and being analyzed by the system, are a multiple phoneme string, containing a number of subgroups, then video and aural indications are provided to the user, indicating the quality or correctness of his pronunciation.

If, however, the system is being operated in the phonemic level mode, a further non-linear transform is used to project the neural network output onto the visual space of the display, so as to provide real time visual feedback, as described above in conjunction with Fig. 3C.

It will be appreciated by persons skilled in the art that the scope of the present invention is not limited by what has been shown and described above, merely by way of example. The scope of the invention is limited, rather, solely by the claims, which follow.

CLAIMS

1. A system for teaching speech pronunciation or reading to a student, which comprises:

a memory for storing a plurality of speech portions;

a playback system, associated with said memory, for indicating to a student a speech portion to be practiced;

an algorithm associated with said memory;

a speech portion selector for selecting a speech task to be practiced; and

a sound recorder connected to said algorithm and operative to sense and record an utterance by a student, and to provide the utterance to said algorithm in signal form,

wherein said algorithm is operative to compare the utterance with the speech task to be practiced and to evaluate the accuracy of the utterance, and is further operatively associated with said playback system, and is operative to cause it to provide to the student, an indication of the accuracy of the utterance.

2. A system according to claim 1, wherein said algorithm is operative to perform speech feature extraction so as to quantify spoken sounds in accordance with predetermined parameters, and is further operative to enhance parameters statistically representative of the selected speech task.

3. A system according to claim 2, wherein said algorithm is further operative, during said speech feature extraction to determine a plurality of speech parameters, including the cepstrum, the 1st and 2nd cepstral derivatives, pitch, energy, zero crossing, or any other parameter or set of parameters derived from the speech signal.

4. A system according to claim 3, wherein said algorithm is further operative to apply predetermined statistical weighting functions to predetermined speech parameters, thereby to enhance said predetermined speech parameters and so as to provide a set of weighted parameters.

5. A system according to claim 4, wherein said algorithm is yet further operative to perform principal component analysis for performing a mathematical transform of said predetermined parameters, thus providing linear combinations of said weighted parameters, and thereby to provide a new set of parameters, ranked in accordance with variability and relevance to the speech task.

6. A system according to claim 5, wherein said algorithm is further operative to truncate said new set of parameters so as to reduce dimensionality thereof, while retaining a predetermined number of statistically representative parameters.

7. A system according to claim 2, wherein said speech portion selector is operative to select a phoneme in a selected phoneme class,

and wherein said algorithm is also operative to perform a cluster analysis so as to determine whether or not a phoneme present in the utterance belongs to the selected phoneme class.

8. A system according to claim 3, wherein said algorithm is further operative to evaluate the accuracy of an uttered phoneme, in accordance with said cluster analysis, by determining the location of the phoneme relative to an associated phonetic space.

9. A system according to claim 8, wherein said algorithm is operative to determine the location of a spoken sound relative to the associated phonetic space, by employing a non-linear, continuous classification network for comparing the occurrence of said predetermined speech parameters with the extracted speech parameters, operative to provide an output corresponding thereto, and further, by projecting the classification network output onto said playback system.

10. A system according to claim 9, wherein said algorithm operates said playback means to again display the inaccurately uttered phoneme for practicing by the student.

11. A system according to claim 10, wherein said sound recorder is operative to sense and record a repeated phoneme, and to provide the repeated phoneme to said algorithm in signal form for evaluation thereof, and wherein said algorithm is operative to evaluate the repeated phoneme for accuracy.

12. A system according to claim 2, wherein said speech portion selector is operative to select a multiple phoneme string,

and wherein said algorithm is also operative to evaluate whether an uttered multiple phoneme string corresponds to the selected multiple phoneme string.

13. A system according to claim 12, wherein said algorithm is operative to identify and evaluate for accuracy subgroups present in the uttered multiple phoneme string.

14. A system according to claim 13, wherein said algorithm is operative to cause said playback system to provide to the student a sensible indication of the inaccurate subgroups in the uttered multiple phoneme string.

15. A system according to claim 13, wherein said algorithm further determines the nature of the inaccuracy of each subgroup and further, in the event that the nature of the inaccuracy is determined to be segmental, to cause said playback means to display each phoneme of the subgroup for practicing by the student.

16. A system according to claim 1, wherein said playback system comprises:

a visual playback system, and

an audio playback system,

and wherein, in response to selection of a selected speech portion by a student, said visual playback system is operative to display a visual image indicating the speech portion selected, and said audio display means is operative to provide an audible indication of the speech portion selected.

17. A system according to claim 16, wherein said playback system is operative to provide in real time a dynamic visual image indicating the accuracy of the utterance.

18. A system according to claim 17, wherein said playback system comprises:

a display of a movable visual image which is movable between first and second locations on said display, wherein said first location is a start location at which said visual image is located prior to sensing of a sound by said recorder, and wherein said second location is a target location, towards which said playback system is operative to move said movable visual image in real time as an indication of the accuracy of the utterance.

19. A system according to claim 18, wherein the distance between said movable visual image and said target location is inversely proportional to the accuracy of the utterance.

20. A system according to claim 1, wherein at least two of said memory, said playback system, and said algorithm are located remotely from one another, and are connected via a communications link.

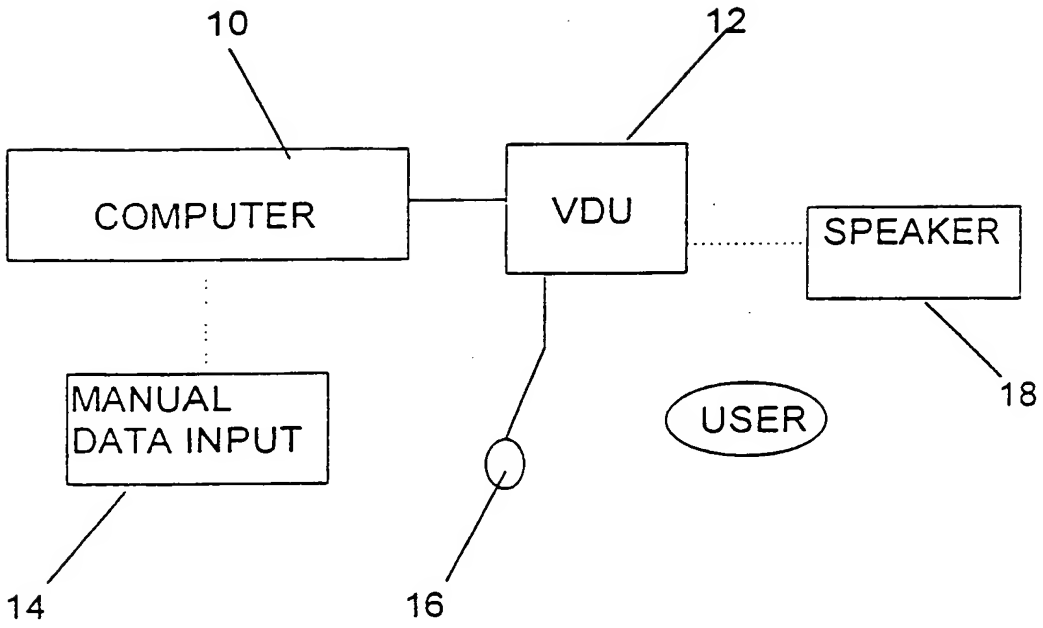


FIG. 1

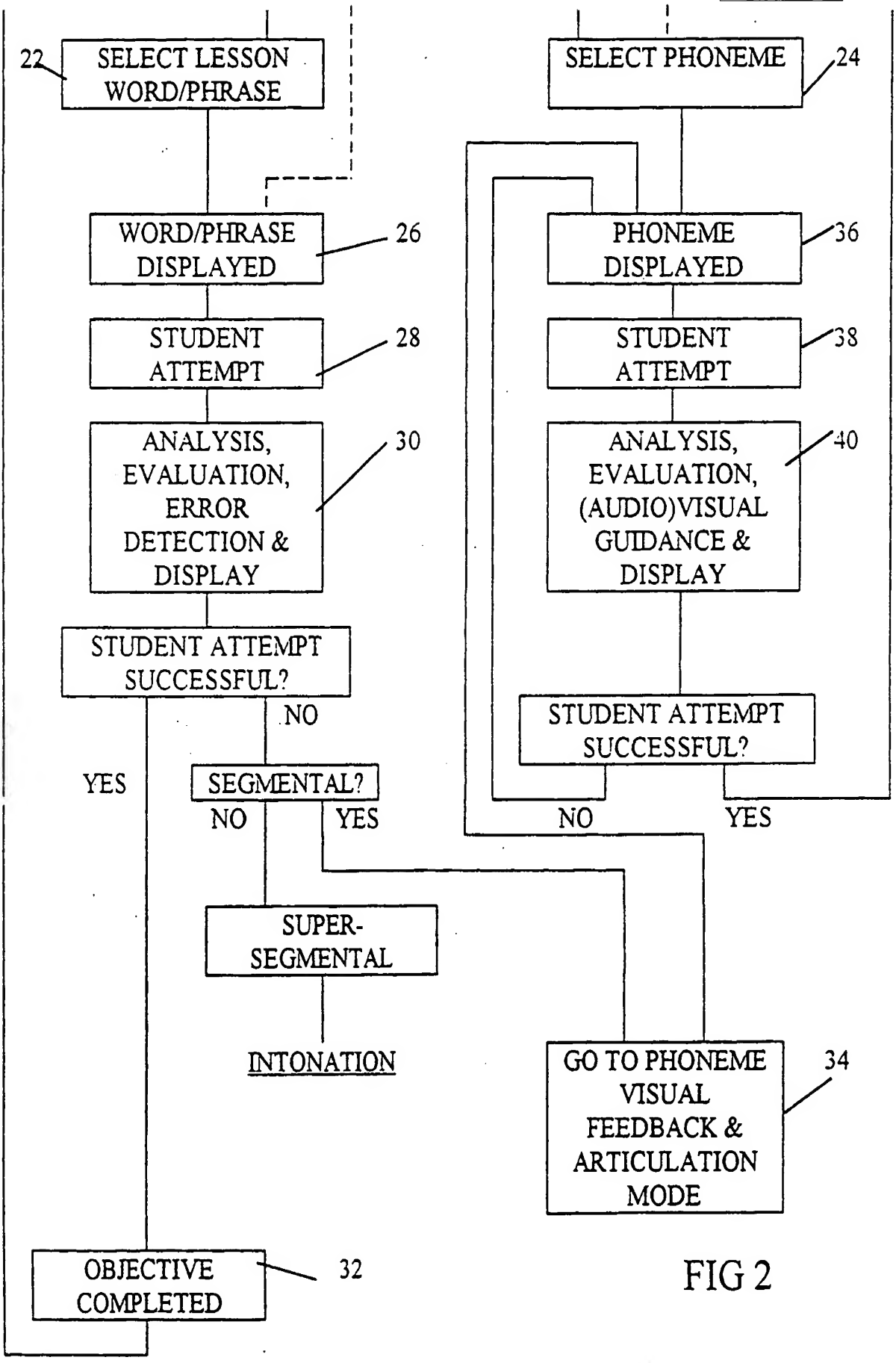


FIG 2

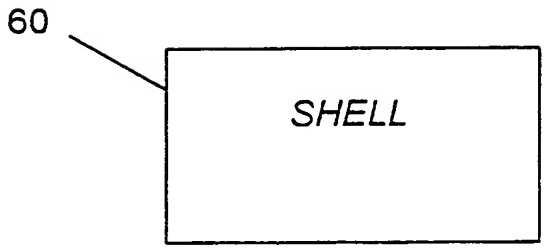


FIG. 3A

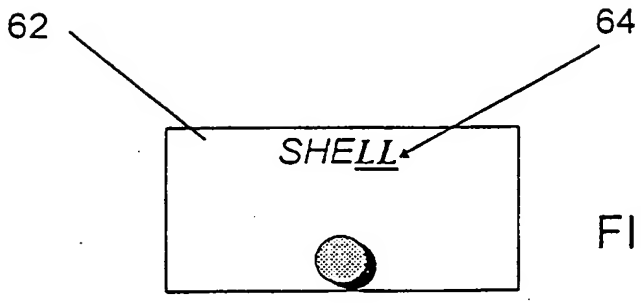


FIG. 3B

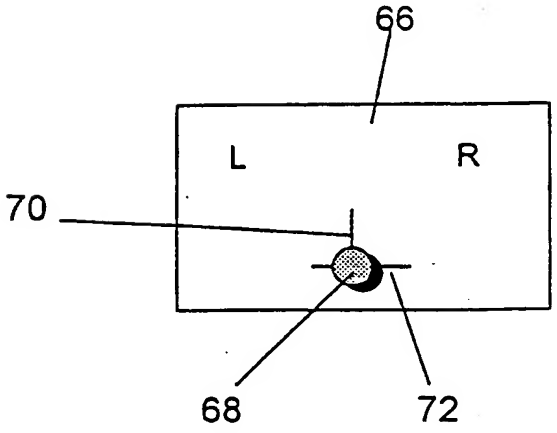


FIG. 3C

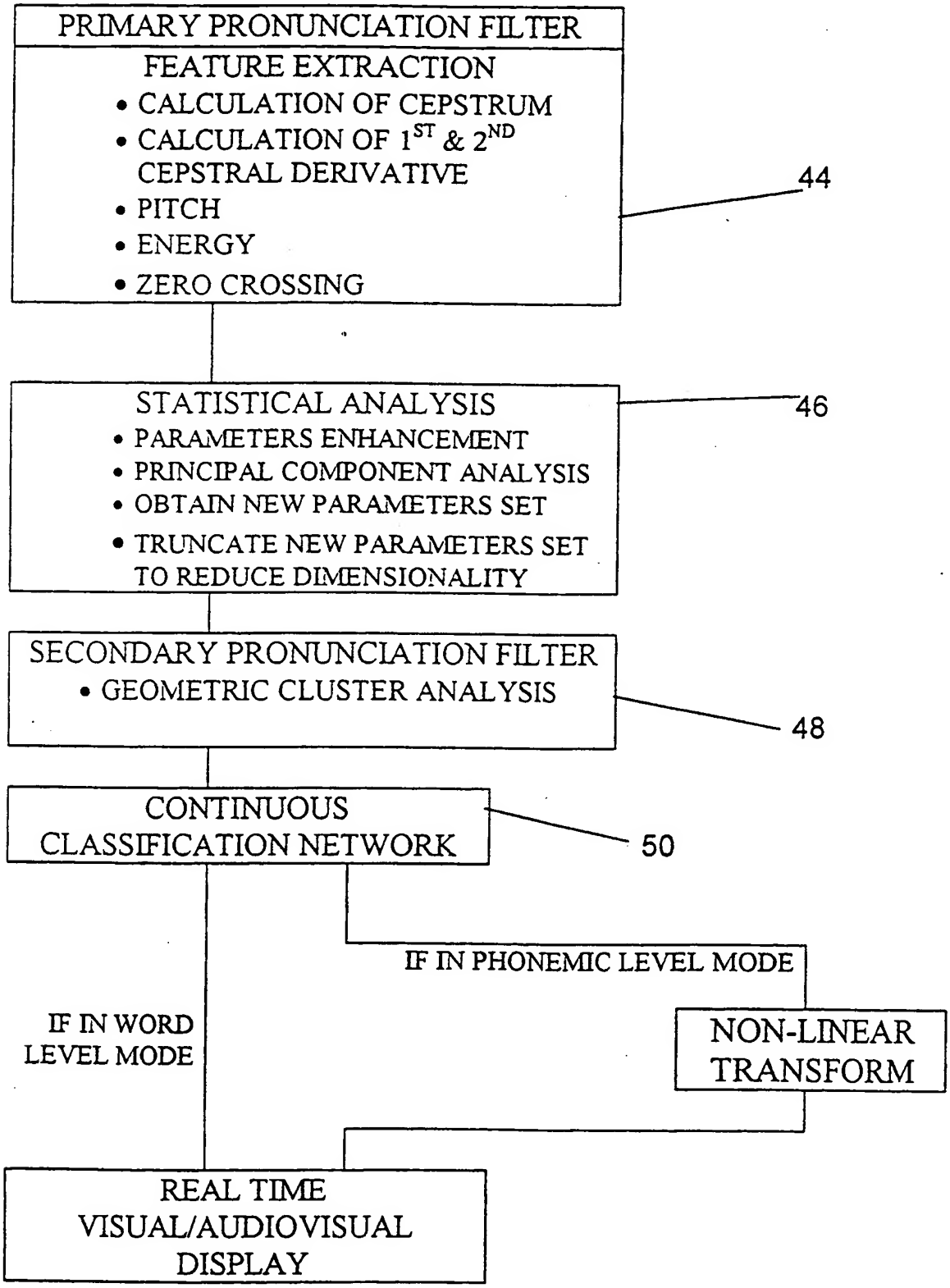


FIG 4

INTERNATIONAL SEARCH REPORT

International application No.
PCT/IL98/00426

A. CLASSIFICATION OF SUBJECT MATTER

IPC(6) :G09B 19/04
US CL :434/185

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 434/118,156,169,185,307R,308,309,323,362,365; 704/1,9,200,211,213,241-231; 381/34.

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

APS
search terms: speech pronunication, utterance, cepstrum, sepstral

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X --- Y	WO 94/17508 A (SHPIRO ET AL) 04 August 1994, see Figs. 1-11.	1,2,7,12-17 --- 3-6,8-11,18 -20
X --- Y	WO 94/10666 A (RUSSELL ET AL) 11 May 1994, see Figs. 1-17.	1,2,7,12-17 --- 3-6,8-11,18 -20
X --- Y	WO 91/00582 A (BOZADJIAN) 10 January 1991, see Figs. 1-3.	1,2,7,12-17 --- 3-6,8-11,18-20

[X] Further documents are listed in the continuation of Box C. [ ] See patent family annex.

*	Special categories of cited documents:	"T"	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A"	document defining the general state of the art which is not considered to be of particular relevance	"X"	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E"	earlier document published on or after the international filing date	"Y"	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L"	document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&"	document member of the same patent family
"O"	document referring to an oral disclosure, use, exhibition or other means		
"P"	document published prior to the international filing date but later than the priority date claimed		

Date of the actual completion of the international search 07 DECEMBER 1998	Date of mailing of the international search report 20 JAN 1999
Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231 Facsimile No. (703) 305-3230	Authorized officer JOE H. CHENG Telephone No. (703) 308-2667

Sheila Venty
Paralegal Specialist
Technology Center 3700

Form PCT/ISA/210 (second sheet)(July 1992)*

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X, E ------- Y, E	US 5,813,862 A (MERZENICH ET AL) 29 September 1998, see Figs. 1-17.	1,2,7,12-17,20 ---------- 3-6,8-11,18,19
Y	US 5,278,942 A (BAHL ET AL) 11 January 1994, see Figs. 1-5.	1-20
X	US 3,881,059 A (STEWART) 29 April, 1975, see Figs. 1-8.	1-20